Journal of Choice Modelling 48 (2023) 100413

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

A discrete choice modeling framework of heterogenous decision rules accounting for non-trading behavior

Evanthia Kazagli [a,*], Matthieu de Lapparent [b]

[a] *School of Management and Engineering Vaud (HEIG-VD), University of Applied Sciences and Arts Western Switzerland (HES-SO), Avenue des sports 20 - CP 521 - 1401 Yverdon-les-Bains, Switzerland*

[b] *HEIG-VD. HES-SO, Avenue des sports 20 - CP 521 - 1401, Yverdon-les-Bains, Switzerland*

A R T I C L E I N F O

Keywords:
Contextual effects
Discrete choice modeling
Non-trading behavior
Probabilistic decision process models
Relative advantage
SP data

A B S T R A C T

We present a discrete choice modeling framework with heterogeneous decision rules accounting for non-trading behavior. The proposed approach builds upon the state-of-the-art probabilistic finite mixture models and tackles non-trading behavior while accounting for inertia effects and serial correlation in the SP data, and contextual effects on the probability of an individual employing a specific decision rule. The framework involves three subpopulations of decision-makers, referred to as respectively as pure utility-maximizers, utility-maximizers with strong preference for one alternative, and non-traders non-utility-maximizers employing a non-trading heuristic. The second subpopulation is expected to exhibit non-trading behavior, despite making trade-offs consistent with utility maximization. Our goal is to disentangle the two types of manifested non-trading behavior. We assume that the manifestation of non-trading behavior – by otherwise utility-maximizing individuals – may be driven by important *context variables*. In order to accommodate this assumption in the modeling framework, we define and add a relative advantage (RA) component in the class-membership model. Finally, we apply the framework to a Swiss stated preferences (SP) mode choice case study, and demonstrate the impact of accounting for non-trading behavior on the value of time estimates.

1. Introduction

Non-trading behavior refers to the case where an individual always chooses the same alternative across choice situations (Hess et al., 2010). This type of behavior is often observed in stated preferences (SP) choice surveys, where respondents are requested to answer several hypothetical choice tasks. Along similar lines, in real life contexts, we encounter *the habitual selection*, with the individual choosing what she chose last time she had to make the same or a similar choice. Opposite to these lies *the variety-seeking selection*, with the individual choosing alternatives that have not been previously chosen. Such decisions-making strategies[1] – allowing the individual to minimize the cognitive effort – are highly pertinent to routine choices (see e.g. Adamowicz and Swait, 2012, in the context of food choices). They belong to the group of suboptimal decision strategies, commonly referred to as heuristics, connoting the omission of part, or all, of the information by the individual in order to make decisions faster and simpler, as opposed to normative decision strategies that assume a rational individual with almost complete information and sufficient capacity to process it for making *trade-offs* that result in an "optimal" choice. A comprehensive review of the decision heuristics within the discrete

* Corresponding author.
E-mail addresses: evanthia.kazagli@heig-vd.ch (E. Kazagli), matthieu.delapparent@heig-vd.ch (M. de Lapparent).

[1] A *decision-making process, strategy or rule* is a mental mechanism through which a person arrives at making a choice. In the remainder of the paper, we use the terms decision-making process, strategy and rule interchangeably, with the latter being mostly used in a choice modeling context.

https://doi.org/10.1016/j.jocm.2023.100413
Received 12 May 2021; Received in revised form 23 January 2023; Accepted 24 April 2023
Available online 10 May 2023

choice modeling (DCM) framework with SP data is presented by Leong and Hensher (2012a). After discussing the contribution of decision heuristics, as well as this of *contextual effects*, in explaining choice behavior, the authors suggest that a logical way forward would be to *"consider the use of mixture models, where multiple heuristics are weighted in a utility function, using weighting functions that depend on the socio-economic characteristics of the respondent and other choice context variables, including individual-specific perceptions data, where available"*.

The nature of the choice to be made may trigger a specific type of strategy. For instance, mode choices are found to be rather habitual (see e.g. Cantillo et al., 2007; Cherchi and Manca, 2011; Cherchi and Cirillo, 2014; Schmid et al., 2019). Gärling and Axhausen (2003) and Cantillo et al. (2007) define habit or inertia in the context of choice behavior as the "reluctance to change"[2]. High inertia may be manifested as non-trading behavior. In this context, non-trading can also be described as lexicographic behavior with respect to the alternative[3]. Indeed, Hess et al. (2010) identify the possible drivers behind non-trading behavior in SP data; these are (i) the strong preference towards a particular alternative, by an otherwise utility-maximizing individual, (ii) the *non-trading heuristic* employed by a non-utility maximizing respondent due to fatigue, boredom, irrelevance of the attribute values, and more, and (iii) some sort of political or strategic behavior, such as never choosing a tolled road alternative. The authors argue that respondents in the first category, i.e. utility maximizers with strong preference towards a specific alternative, should not be excluded from a utility maximizing model, while those in the other two categories should ideally be identified and excluded from the model estimation in order to avoid biases in the derivation of policy indicators, such as the willingness to pay. They acknowledge the fact though that, in the majority of cases, it is not possible to discriminate between the two types of non-trading behavior.

The importance of identifying and modeling non-trading behavior in random utility frameworks, as pointed out by Hess et al. (2010), and the significance of *context* in the relevance of a decision rule, as pointed out by Leong and Hensher (2012a) and Hensher (2019), have motivated the modeling framework proposed in this paper. The work concerns a practical application that focuses on a model specification that accounts for contextual effects, which are based on objective and measurable factors, in order to tackle non-trading behavior. The proposed approach involves three subpopulations of decision-makers, referred to respectively as (i) pure utility-maximizers, (ii) utility-maximizers with strong preference for a specific alternative, and (iii) non-traders non-utility-maximizers (pure non-traders), adopting the non-trading heuristic to minimize the effort, for instance. It postulates that the manifestation of non-trading behavior, by otherwise utility-maximizing individuals, may be driven by important context variables, and more specifically by the overall relative advantage with respect to these variables, of ones preferred alternative over the remaining alternatives in the choice task. As discussed by Hess et al. (2010), in the case of extreme preferences, the choice design "*may not be able to offer the respondent sufficiently attractive alternatives to their preferred mode*". In order to test this, we include a relative advantage (RA) component (Leong and Hensher, 2014) in the class-membership model (CMM). Leong and Hensher (2014) propose the relative advantage maximization model (RAM) as an extension to the conventional linear additive random utility model (RUM) in order to account for context dependency in the representation of the choice set, and subsequently the impact of constructed preferences on the choice of an alternative. The use of the RA component here is different from the one in the RAM, in that it adds the RA component in the CMM, rather than in the class-specific choice model (CSM). Yet, its notion and formulation are the same as in Leong and Hensher (2014). Consequently, the RA component is used to describe the probability of an individual belonging to the second class, rather than her probability of choosing an alternative. This model specification aims at disentangling the two types of the manifested non-trading behavior.

Methodologically, the framework is based on the well-established probabilistic decision process modeling (PDPM) approach[4]. Such approaches have been presented by various papers in the general DCM literature tackling heuristics and non-compensatory choice behaviors. Different mixtures of decision rules have been considered by e.g. Elrod et al. (2004) who proposed an integrated model of disjunctive/conjunctive screening rules; Hensher and Greene (2010) who analyzed attribute non-attendance and dual processing with a latent class specification; Zhu and Timmermans (2010) who assume context-dependent preferences and decision heuristics and incorporate conjunctive, disjunctive and lexicographic decision rules in their modeling framework; Hess et al. (2012) who applied four different mixture of the RUM with other decision rules (lexicography, heterogeneous reference points, elimination by aspects and random regret minimization) in four case studies; Leong and Hensher (2012b) who have performed an exploratory analysis of multiple mixtures of heuristics, including the reference point revision and the majority of confirming dimensions; McNair et al. (2012) who presented a PDPM tackling the value learning and the strategic misrepresentation heuristics; Hess and Stathopoulos (2013) who present a mixture of random utility and random regret model and use a latent variable modeling approach in analyzing the probability of each decision rule being followed by a respondent; Hensher et al. (2013) who apply a random parameter PDPM with different levels of attribute attendance (full/non-attendance, as well as aggregation of common attributes) to a SP for car commuters; Boeri et al. (2014) who analyze the performance of a utility maximization-regret minimization mixture in the context of traffic calming schemes, to find that unfamiliar users are more likely to be regret minimizers in the SP, in comparison with familiar users; Balbontin et al. (2017) and Hensher et al. (2018) who tackle the heterogeneity in decision processes in the presence of risk; Dey et al. (2018) apply a utility maximization-regret minimization mixture to bicycle route choice; and more.

This is to our knowledge the first work to integrate contextual effects in the weighting functions that define the class-membership model and the first to tackle non-trading behavior within the random utility framework. We present an application of the approach

[2] Inertia is linked to the concept of *state dependence* in the context of choice dynamics, which describes the effect that an already made choice has on the likelihood of a choice in future occasions (McAlister et al., 1991).

[3] The result of the application of this lexicographic decision rule is that respondents do not make trade-offs among the various attributes.

[4] This is essentially a latent class modeling approach, where each class is characterized by different preference measures, as a result of the differences in the underlying decision-making process.

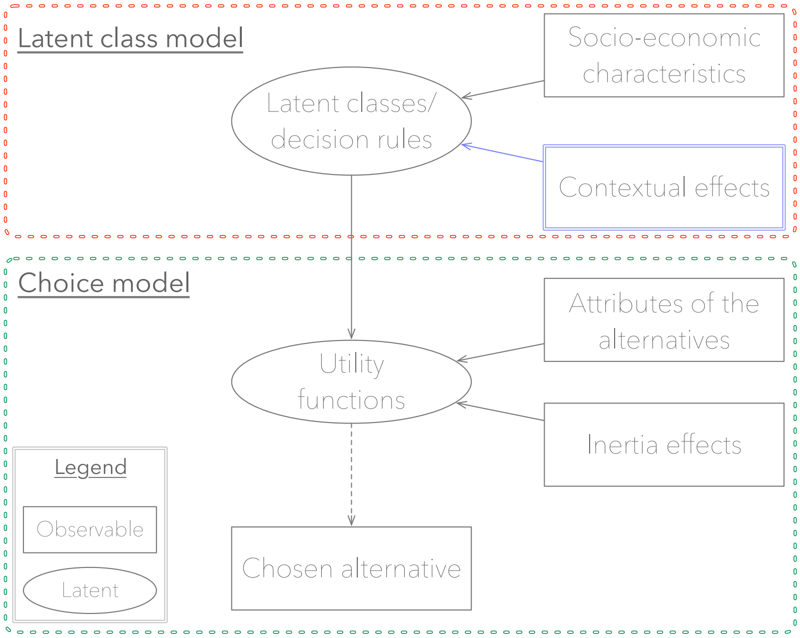

Fig. 1. Illustration of the conceptual modeling framework.

to a Swiss stated preferences (SP) mode choice dataset and demonstrate the influence of tackling non-trading behavior on the resulting value of time estimates. The model takes the form of a mixed logit. Its estimation treats serial correlation, that is the dependence of the responses provided by the same individual, through the inclusion of respondent-, alternative- and class-specific error components.

The remainder of the paper is organized as follows. Section 2 delineates the modeling framework. Section 3 presents the Swiss mode choice case study. Section 4 presents the results from the application of the proposed modeling framework to the available data. Section 5 summarizes the finding of the work and identifies directions for further research

2. Modeling framework

We consider three classes of individuals, namely (i) the utility-maximizers, (ii) the utility-maximizers with strong preference for a specific alternative, and (iii) the non-traders non-utility-maximizers employing the effort-reduction non-trading heuristic (NTH). We expect a utility maximizer with a strong preference for a specific mode to exhibit non-trading behavior, as she is likely to persistently choose it, unless, possibly, another alternative in the choice context is notably more attractive.[5] More specifically, we assume that the persistence of choosing a specific mode may not be merely inherent – e.g., not merely due to habit – but likely to be triggered by the context, with the latter being characterized by the attractiveness the preferred alternative against the remaining alternatives. This tenacity is distinct from the effort-reduction NTH.

The goal of the approach delineated in this section is to discriminate between utility-maximizers and non-traders.

2.1. Model formulation

The framework builds upon the state-of-the-art finite mixture models, under the assumption that each individual in the data is making choices based on a specific, yet unknown, decision rule. This assumption gives rise to a probabilistic decision process modeling (PDPM), or latent class, approach (see e.g. Hess et al., 2012; McNair et al., 2012). The conceptual framework is depicted in Fig. 1.

The probability that an individual n choses alternative i given the choice set of alternatives C_n and the set of possible decision rules $\mathcal{D}$ is defined as

$$\mathrm{P}_n(i \mid C_n) = \sum_{d \in \mathcal{D}} \mathrm{Pr}_n(i \mid C_n; d) \cdot \mathrm{Pr}_n(d), \tag{1}$$

[5] This is a common issue in SP designs, where the attributes of the alternatives in the choice experiment are often pivoted around the reference alternative, where the reference alternative may correspond to the last chosen alternative or the most commonly chosen alternative, in a real context.

where $\Pr_n(i \mid C_n; d)$ is the probability that n chooses i given that she uses decision rule d; this is the class-specific choice model (CSM). Its specification depends on the assumption about the decision rule. $\Pr_n(d)$ is the probability that n adopts decision rule d to arrive at making a choice—or else the probability she belongs to class d[6]. $\Pr_n(d)$ is the class-membership model (CMM). It can be modeled as a function of the decision-maker's characteristics, choice context variables, as well as (depending on availability) individual-specific attitudinal/perceptual data (see e.g. Hess and Stathopoulos, 2013). In what follows we define the CSMs and the CMM for the proposed framework.

2.2. Model specification

For the modeling framework depicted in Fig. 1 to be operational, we need to define

1. the *class-specific model* (CSM), describing the probability $\Pr_n(i \mid C_n; d)$;
2. the *class-membership model* (CMM), describing the probability $\Pr_n(d)$

The reluctance to change from a specific alternative is captured through the inclusion of inertia effects in the CSM, while the effect of the context on the persistence of choosing the alternative is incorporated in the CMM and represented by the RA component. For the specification of the model on SP data, we introduce the notation $t \in T$, denoting the choice task in the sequence of choices tasks T made by n. We begin with the definition of the CSM for each class of respondents.

2.2.1. Definition of the CSM

Pure random utility maximizers For the two classes of utility maximizers, consistent with the RUM framework, the probability $\mathrm{P}_{nt}^{\mathrm{RUM}}(i \mid C_n;\, \mathrm{RUM})$ that an individual n chooses alternative i, with associated utility U_{in}^t in choice task t, from her choice set C_n, is equal to the probability of U_{in}^t being the highest utility among all alternatives in C_n. This can be formally expressed as

$$\mathrm{P}_{nt}(i \mid C_n;\, \mathrm{RUM}) = \Pr(U_{in}^t \geq U_{jn}^t, \forall j \in C_n), \tag{2}$$

where U_{in}^t is decomposed in the following manner

$$U_{in}^t = V_{in}^t + \varepsilon_{in}^t, \tag{3}$$

with V_{in}^t being the deterministic or systematic (observable) component of the utility and ε_{in}^t being the random or stochastic (unobservable) component (to be defined according to the choice context). The systematic part of the utility for the *first class* is

$$V_{in}^{1t} = \beta_{i0}^1 + \sum_k \beta_{ik} x_{ink}^t, \tag{4}$$

where β_{i0}^1 is an alternative specific constant and β_{ik} is the influence of the kth attribute x_{ink}^t on the utility of the alternative i.

Random utility maximizers with strong preference for a specific alternative The systematic utilities of the *second class* are exactly the same as before, with the addition of the inertia effects capturing the inherent preference of n for a specific i

$$V_{in}^{2t} = \beta_{i0}^2 + \sum_k \beta_{ik} x_{ink}^t + v_i \times I_{in}, \tag{5}$$

where v_i are the inertia parameters and I_{in} are alternative-specific inertia dummy variables, being equal to one, if i corresponds to the preferred i of n, and zero otherwise (as in Bradley et al., 1996). We assume that, apart from the inertia effects, the same utility functions hold for all utility-maximizers (first and second class). We postulate class- and alternative-specific constants β_{i0}^d, while whether the attribute sensitivities β_{ik} are class-specific or not is determined during the specification testing.

For the two first classes, we specify the error terms ε_{in}^t for the longitudinal data (similar to e.g. Cantillo et al., 2007)

$$\varepsilon_{in}^{dt} = \eta_{in}^d + \xi_{in}^{dt}, \tag{6}$$

where η_{in}^d are individual-, alternative- and class-specific random variables, invariant over choice tasks $t \in T$, capturing serial correlation due to multiple responses provided by the same individual and ξ_{in}^{dt} are random white noise terms. We specify the serial correlation term as $\eta_{in}^d = \sigma_i \omega_{in}^d$, with the error components ω_{in}^d assumed to be standard normally distributed and σ_{in}^d denoting the standard deviation of η_{in}^d to be estimated. Finally, we assume ξ_{in}^{dt} to be independently and identically distributed (i.i.d.) Extreme Value, allowing us to define $\mathrm{P}_{nt}^{\mathrm{RUM}}(i \mid C_n)$ as

$$\mathrm{P}_{nt}(i \mid C_n;\, \mathrm{RUM}) = \frac{e^{V_{in}^{dt} + \eta_{in}^d}}{\sum_{j \in C_n} e^{V_{jn}^{dt} + \eta_{jn}^d}}, \ \forall i \in C_n. \tag{7}$$

[6] Throughout the document, d denotes both the decision rule and the corresponding class of decision makers, with $d = \{1$: utility-maximizers, 2 : utility-maximizers with strong preference for a specific alternative, 3 : non-traders$\}$, for the sake of simplicity of the notation.

Non-traders For the third class of decision-makers, i.e. for the *non-traders* who use the NTH, we define the heuristic mathematically on the basis of the RUM. The utility function for the NTH is defined deterministically as

$$V_{in}^{3t} = \begin{cases} 0 & \text{if } i \text{ is the preferred alternative } p \text{ of } n, \text{ and} \\ -\infty & \text{otherwise.} \end{cases} \tag{8}$$

and subsequently, by substituting (8) in (7), the choice probabilities become

$$\mathrm{P}_{nt}(i \mid C_n;\ \mathrm{NTH}) = \begin{cases} 1 & \text{if } i \text{ is the preferred alternative } p \text{ of } n, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

2.2.2. Definition of the CMM

The definition of the CMM is based on class membership functions describing the behavioral types of the decision-makers and consequently the probability of a decision-maker in the sample belonging to a class. The class-membership functions of the three classes read as follows

$$F_{1nt} = CSC_1 + \sum_z \beta_z z_n, \tag{10}$$

$$F_{2nt} = CSC_2 + \sum_z \beta_z z_n + \theta \sum_{j \neq p} RA_{nt}(p, j), \tag{11}$$

$$F_{3nt} = CSC_3, \tag{12}$$

where CSC_d are class-specific constants, z_n represents the socioeconomic characteristics of the respondent and θ is the parameter associated with the RA component RA_{nt} in choice task $t \in T$[7]. For the definition of the RA_{nt} we adopt the formulation described by Leong and Hensher (2014)[8]. We define the relative advantage RA_{nt} of the preferred alternative p with respect to every other alternative $j \neq p$ in the choice task t as

$$RA_{nt}(p, j) = \frac{A_{nt}(p, j)}{A_{nt}(p, j) + D_{nt}(p, j)}, \tag{13}$$

where $A(p, j) = \sum_k A_k(p, j)$ and $D(p, j) = \sum_k D_k(p, j)$ are, respectively, the overall advantage and disadvantage of p over j over all relevant attributes k. The advantage of p over j with respect to k is defined as $A_k(p, j) = D_k(j, p) = \ln[1 + \exp(\beta_{pk} X_{pk} - \beta_{jk} X_{jk})]$, if $v_k(X_{pk}) \geq v_k(X_{jk})$, and zero otherwise, with $v_k(X_{jk})$ being the utility of attribute k for alternative j. Finally, the overall RA of p over all $j \neq p$ is $\sum_j RA(p, j)$.

For the application of the framework to SP data, it is important to bear in mind that – contrary to the common approach – the CMM probability $\Pr_{nt}(d)$ is not constant across choice tasks for a given individual n. This has implications for the definition of the unconditional probability of the choice sequence $y_{n1}, \ldots, y_{nT}$. When $\Pr_{nt}(d) = \Pr_n(d),\ \forall\, t \in T$, the unconditional probability of the sequence of choices is derived by taking the expectation over all $d \in D$

$$\Pr_n\left(y_{n1}, \ldots, y_{nT}\right) = \sum_{d \in D} \Pr_n(d) \prod_{t=1}^{T} \Pr_{nt}\left(y_{nt} | d\right), \tag{14}$$

where y_{nt} is equal to one if alternative i is chosen by individual n in choice task t, and zero otherwise. When the class-membership varies across choice tasks, the unconditional probability of the choice sequence $y_{n1}, \ldots, y_{nT}$ becomes

$$\begin{aligned}
\Pr_n\left(y_{n1}, \ldots, y_{nT}\right) &= \sum_{d \in D}^{1} \cdots \sum_{d \in D}^{T} \Pr_{nt}\left(y_{n1}, \ldots, y_{nT}, d\right), \\
&= \sum_{d \in D}^{1} \cdots \sum_{d \in D}^{T} \Pr_{nt}(d) \cdot \Pr_{nt}\left(y_{n1}, \ldots, y_{nT} | d\right), \\
&= \sum_{d \in D}^{1} \cdots \sum_{d \in D}^{T} \Pr_{n1}(d) \cdots \Pr_{nT}(d) \cdot \Pr_n\left(y_{n1}, \ldots, y_{nT} | d\right), \\
&= \sum_{d \in D}^{1} \cdots \sum_{d \in D}^{T} \Pr_{n1}(d) \cdots \Pr_{nT}(d) \cdot \Pr_n\left(y_{n1} | d\right) \cdots \Pr_{nT}\left(y_{nT} | d\right), \\
&= \sum_{d \in D}^{1} \cdots \sum_{d \in D}^{T} \Pr_{n1}(d) \Pr_{n1}\left(y_{n1} | d\right) \cdots \Pr_{nT}(d) \Pr_{nT}\left(y_{nT} | d\right), \\
&= \sum_{d \in D}^{1} \Pr_{n1}(d) \Pr_{n1}\left(y_{n1} | d\right) \cdots \sum_{d \in D} \Pr_{nT}(d) \Pr_{nT}\left(y_{nT} | d\right), \\
&= \prod_{t=1}^{T} \left(\sum_{d \in D} \Pr_{nt}(d) \cdot \Pr_{nt}\left(y_{nt} | d\right) \right).
\end{aligned} \tag{15}$$

[7] Recall that this is contrary to the traditional use of the RA model, where the RA component is included in the utility functions of the alternatives to capture the context dependence of preferences. Here, we evaluate the influence of the context on the choice of a decision rule.

[8] Earlier formulations of the RA model can be found in Tversky and Simonson (1993) and Kivetz et al. (2004)

Table 1
Availability of modes.
Source: (Adapted from Weis et al. (2021)).

Car available	Trip length[a]	Reported mode	Available modes
No	Short	Walk	Walk/bike/PT
No	Short	Bike	Walk/bike/PT
No	Short	PT	Walk/bike/PT
No	Long	PT	–
Yes	Short	Walk	Walk/car/PT
Yes	Short	Bike	Bike/car/PT
Yes	Short	PT	Walk or bike/car/PT
Yes	Short	Car	Walk or bike/car/PT
Yes	Long	PT	Car/PT
Yes	Long	Car	Car/PT

[a]Where short trips are those under 10 km.

We can further specify respondent- and class-specific random variables $\eta_{dn} = \sigma_d \omega_{dn}$, in order to capture the correlation in the class-membership across observations of the same individual. Same as for the CSMs, the error components ω_{dn} are standard normally distributed and the standard deviation σ_d of η_{dn} is to be estimated. Assuming once again i.i.d. Extreme Value random white noise error terms f^t_{nd}, the class membership probabilities can be defined as

$$\Pr_{nt}(d) = \frac{e^{F^t_{dn}+\eta_{dn}}}{\sum_{c\in D} e^{F^t_{cn}+\eta_{cn}}}, \forall d \in \mathcal{D}. \tag{16}$$

Finally, it is important to note that the model is blind as to whether a respondent's choice sequence in the SP experiment exhibits non-trading behavior or not. Every n is eligible to belong to every d with probability $\Pr_{nt}(d)$.

2.3. Model estimation

We estimate mixed logit models with random error components (Sections 2.2.1–2.2.2) distributed across individuals but constant across observations provided by the same individual. The likelihood function of the sample, integrated over the domain of η^d_{in}, η_{dn}, can be expressed as follows

$$\mathcal{L} = \prod_n \int_{\eta^d_{in},\eta_{dn}} \prod_{t=1}^{T} \sum_d \Pr_{nt}(d \mid z_n, RA_{nt}, \eta_{dn}) \Pr_{nt}(y_{nt} = 1 \mid x^t_{ink}, I_{in}, \eta^d_{in}; d) f(\eta^d_{in}) h(\eta_{dn}) d\eta^d_{in} d\eta_{dn}. \tag{17}$$

As Eq. (17) does not have a closed form for the probabilities, it has to be estimated using maximum simulated likelihood techniques (see Train, 2009).

3. Case study

This section presents the Swiss SP mode choice dataset that is used for the illustration and empirical application of the framework.

3.1. Data

We use data from a stated preference (SP) survey for mode choice behavior that was conducted in Switzerland in 2015[9]. The SP design is based on variations in the attribute values that are centered around each respondent's RP trip characteristics. In total, four modes appear in the experiments: (i) walking, (ii) bike, (iii) car and (iv) public transport. Each respondent was presented with a choice set of two-three alternatives, depending on her availability of transport means and the length of her reported (last) trip for a specific trip purpose (Table 1). Public transportation is the only alternative that is always in the choice set, while walk, bike and car are available for 22.5%, 33.7% and 94.1% of the respondents, respectively. We refer to the chosen mode with which the real reported trip was conducted as the respondent's revealed preference (RP) choice/mode. The data about the RP choice for the trip in question is available, along with the socio-economic characteristics of the respondent and her indications about which attributes of the alternatives she attended for making a choice.

The sample includes 1522 respondents generating 1522×8 choice tasks $=$ 12176 observations—after excluding (i) the observations from the pre-tests, (ii) respondents who did not report their household income and (iii) those who did not answer all 8 experiments in the design. Table 2 presents the descriptive statistics of the attributes of the alternatives in the survey. Concerning the characteristics of the sample, 813 respondents are men and 709 women. The average age of the sample is 48 years old, with a minimum age of 18 and a maximum age of 86. 1423 respondents have a driving license and 913 own some type of travel card.[10]

[9] *Data source*: Stated preferences surveys for transport behavior 2015, Federal Office for Spatial Development ARE, Bern, 2017, http://www.are.admin.ch/statedpreference. We refer the reader to Weis et al. (2021) for more details regarding the survey design and the dataset.
[10] E.g. a country-wide pass or a half-fare card.

Table 2
Descriptive statistics of the mode attributes in the survey.

Total travel time	Available in # choice tasks	min	max	mean	stdev
Walking	2736	6.0	467.0	51.9	41.7
Bike	4104	2.0	140.0	22.3	15.0
car	11 456	2.0	395.0	29.5	28.8
PT	12 176	2.0	1004.0	43.9	44.9
Total travel cost	**Available in # choice tasks**	**min**	**max**	**mean**	**stdev**
Car	11 456	0.9	108.7	7.5	7.2
Public transport	12 176	1.4	125.0	7.3	10.4
# of transfers	**Available in # choice tasks**	**min**	**max**	**mean**	**stdev**
PT	12 176	0.0	8.0	1.2	1.3

Table 3
Distribution of non-trading respondents over the modes.

Mode	Non-trading respondents per mode	RP mode shares	% non-trading over RP
Walking	45 (5.4%)	94 (6.2%)	47.9%
Bike	87 (10.5%)	129 (8.5%)	67.4%
Car	597 (71.8%)	1033 (67.9%)	57.8%
Public transport	103 (12.4%)	266 (17.5%)	38.7%
Total	832	1522	

Table 4
SP mode shares.

Mode	SP mode shares trading respondents	SP mode shares overall	Mode availability for # of respondents
Walking	35.9 (5.2%)	80.9 (5.3%)	342 (22.5%)
Bike	82.4 (11.9%)	169.4 (11.1%)	513 (33.7%)
Car	325.3 (47.1%)	922.3 (60.6%)	1432 (94.1%)
Public transport	246.5 (35.7%)	349.5 (23.0%)	1522 (100.0%)
Total	690	1522	

Table 5
Statistics on the persistence of the *trading* respondents in choosing their RP choice over the experiments.

Mode	Average persistence	Persistence ≥ 0.75 (6–7 tasks)
Walking	0.401	22.4%
Bike	0.577	52.4%
Car	0.561	41.3%
Public transport	0.514	30.7%
Overall	0.539	38.1%

3.2. Non-trading and context-based inertia evidence

Approximately 55% of the retained respondents, i.e. 832, systematically chose their RP choice across all 8 tasks (Table 3). In Tables 3–4, we denote these individuals as non-trading respondents, because their choices in the experiment exhibit non-trading behavior. Yet, as discussed in the previous sections, these respondents do not necessarily employ the non-trading heuristic. They may as well be utility maximizers characterized by strong inertia towards their habitual mode. Here, we deal with SP data where the context is defined by a RP trip. Hence, inertia is to be interpreted with respect to this reference trip and not as general behavior of the traveler.

We observe that for each RP mode share, the percent of the individuals that fully stick, across all 8 tasks, to their RP choices ranges from 38.7%, for public transport, to 67.4% for bike, with walking and car non-traders reaching respectively 47.9% and 57.8%. For each individual whose choices *do not* exhibit full non-trading behavior, we compute the level of persistence of choosing her RP choice across the SP tasks. The persistence is computed as the ratio of the number of times an individual chooses her RP choice over 8, the total number of tasks in the experiment. That is, if she chooses her RP choice four times out of the 8 experiments, her persistence is 50%. Obviously, for individuals exhibiting non-trading behavior the persistence is equal to one (or 100%).

Table 5 presents statistics regarding the persistence of the respondents depending on their chosen mode of transport in the RP data. Same as for non-trading respondents, among the trading respondents, those who chose bike in the RP data appear to be the more tenacious in choosing their preferred mode, followed by those who chose car, public transport and walking. Overall, the average persistence in the sample of trading respondents is 0.539. 77 of them never chose their RP mode (5.1%) and 22 of them only chose their RP choice once (1.4%). Fig. 2 shows the distribution of the persistence for the sample of trading respondents.

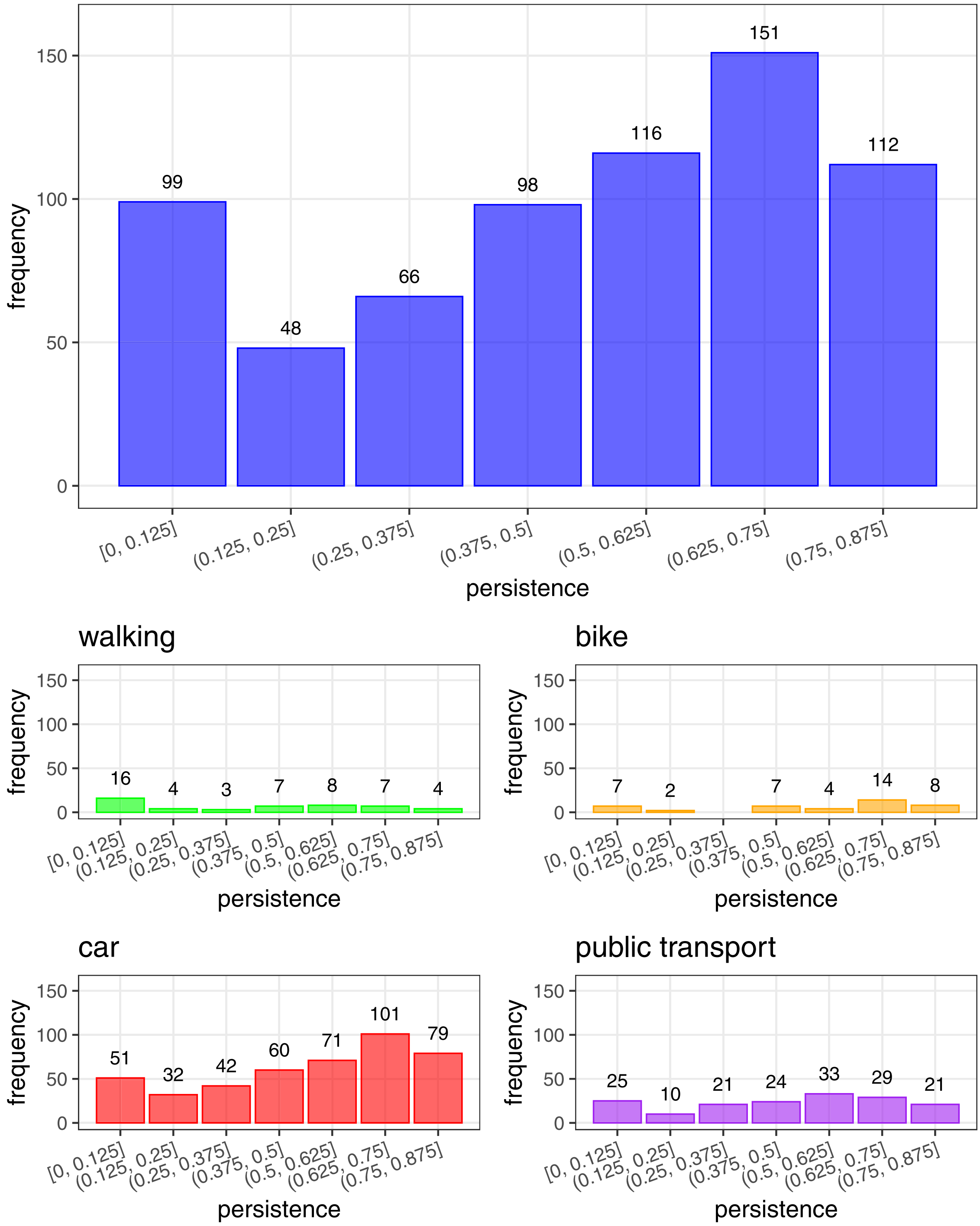

Fig. 2. Persistence of the *trading* respondents in choosing their RP mode over the experiments.

4. Application

In this section, we apply the modeling framework described in Section 2 to the case study presented in Section 3.

4.1. Specification of the CSM

We postulate that – apart from the inertia effects – the same utility functions hold for all utility-maximizers (first and second class). We specify class- and alternative-specific constants ASC_i^d, while after specification testing, we conclude to (i) generic attribute sensitivities for the two classes: $\beta_{ik}^1 = \beta_{ik}^2$, $\forall k$ and (ii) generic time and cost sensitivities for car and public transport $\beta_{\text{totalCost, car}} = \beta_{\text{totalCost, PT}} = \beta_{\text{totalCost}}$ and $\beta_{\text{totalTime, car}} = \beta_{\text{totalTime, PT}} = \beta_{\text{totalTime}}$. That is, we assume that the only driver for belonging to the first or second class is the RA component. One might consider the direct sensitivity to the travel attributes to be different between the two classes—the global sensitivity being different through the RA component. We tested for both class-specific, as well alternative-specific coefficients, but we faced numerical issues.

For the class of *utility maximizers*, the systematic utility functions of the four alternatives are as follows

$$V_{\text{walk}}^1 = ASC_{\text{WALK}}^1 + \beta_{\text{walkTIME}} walkingTime + \beta_{\text{walkTIME}}^{\leq 30\text{min}} walkingTime^{\leq 30\text{min}}, \tag{18}$$

$$V_{\text{bike}}^1 = ASC_{\text{BIKE}}^1 + \beta_{\text{cycleTIME}} cyclingTime, \tag{19}$$

$$V_{\text{car}}^1 = ASC_{\text{CAR}}^1 + \beta_{\text{totalCost}} totalCarCost + \beta_{\text{totalTime}} totalCarTime, \tag{20}$$

$$V_{\text{PT}}^1 = ASC_{\text{PT}}^1 + \beta_{\text{totalCost}} ticketCost + \beta_{\text{totalTime}} totalPTTime + \beta_{\text{numTranfers}}^{\geq 2} numTransfers^{\geq 2}. \tag{21}$$

For *utility maximizers with strong preferences* towards a specific alternative, the systematic utility functions are extended to account for inertia as follows

$$V_{\text{walk}}^2 = ASC_{\text{WALK}}^2 + \cdots + \nu_{\text{walker}} \times I_{\text{walk}}, \tag{22}$$

$$V_{\text{bike}}^2 = ASC_{\text{BIKE}}^2 + \cdots + \nu_{\text{cyclist}} \times I_{\text{bike}}, \tag{23}$$

$$V_{\text{car}}^2 = ASC_{\text{CAR}}^2 + \cdots + \nu_{\text{driver}} \times carAlwaysAvailable \times I_{\text{car}}, \tag{24}$$

$$V_{\text{PT}}^2 = ASC_{\text{PT}}^2 + \cdots + \nu_{\text{PT user}} \times TravelCard \times I_{\text{PT}}, \tag{25}$$

where ν_{walker}, ν_{cyclist}, ν_{driver}, $\nu_{\text{PT user}}$ are the inertia parameters associated with each mode and I_i are the mode-specific inertia dummy variables, being equal to one, if the mode corresponds to the RP choice, and zero otherwise. Unlike previous studies (see eg. Cherchi and Manca (2011) and Schmid et al. (2019)), the inertia variable here is equal to one no matter whether the alternative that corresponds to the RP choice is chosen in the SP task or not. This specification of inertia is considered more appropriate for the present multi-class context, where the objective is to disentangle non-trading behavior due to strong preferences and distinguish it from the effort-reduction NTH. In addition, it avoids endogeneity issues arising from the inclusion of the choice variable in the utility function. Finally, $carAlwaysAvailable$ and $TravelCard$ are dummy variables determining full accessibility to car and possession of public transport subscriptions, respectively. At this point, it is important to keep in mind that some a priori knowledge of an individual's preferred alternative is required, i.e. here the RP choice, so that this does not have to be inferred from the SP data—resulting in endogeneity[11].

For *non-traders non-utility-maximizers* the utility functions are deterministically given by Eq. (8) and depend solely on the RP choice. The specification of the CSMs given by Eqs. (18)–(25) postulates that the two classes of utility maximizers have the same value of travel time savings (VTTS), while for the third class the VTTS is obviously zero.

4.2. Specification of the CMM

The specification testing for the inclusion of socioeconomic characteristics in the class-membership functions did not result in any significant effects. Eventually, only the class-specific constants and the RA component are maintained.

$$F_1 = CSC_1, \tag{26}$$

$$F_2 = CSC_2 + \theta \sum_{j \neq p} RA_{nt}(p, j), \tag{27}$$

$$F_3 = CSC_3. \tag{28}$$

The RA component in this study is computed based on the *total cost* and the *total time* of the alternatives while assuming generic parameters, that is $\beta_{pk} = \beta_{jk} = \beta_k$ and subsequently $A_k(p,j) = D_k(j,p) = \ln[1 + \exp(\beta_k(X_{pk} - X_{jk}))]$, if $\upsilon_k(X_{pk}) \geq \upsilon_k(X_{jk})$, and zero otherwise. Furthermore, these parameters are constrained to be equal to the $\beta_{\text{totalCost}}$ and $\beta_{\text{totalTime}}$ parameters of the CSMs defined in Section 4.1.

[11] We would like to thank an anonymous reviewer for pointing out the importance of highlighting this aspect of the model specification.

Table 6
Summary of goodness of fit.

	BM	PDPM 1	PDPM 2
# of draws	500	500	500
# of parameters	16	19	27
# of respondents	1522	1522	1522
# of observations	12 176	12 176	12 176
$\mathcal{L}(\hat{\beta})$	−4317.45	−4245.11	−4143.47
BIC	8752.14	8629.45	8484.79
AIC	8666.90	8528.22	8340.94

4.3. Estimation results and discussion

We report results from 3 models, gradually extending to the full model specification presented in Sections 4.1–4.2.

- Base model with inertia effects (BM): all respondents are assumed to be utility-maximizers with utility functions given by Eqs. (22)–(25)[12];
- PDPM 1: respondents may belong to each of the following two latent classes: (i) utility-maximizers described by Eqs. (22)–(25) or (ii) pure non-traders described by Eqs. (8); CMM based on Eqs. (27)–(28)[13];
- PDPM 2: full specification as described in Sections 4.1–4.2: (i) utility-maximizers described by Eqs. (18)–(21), (ii) utility-maximizers with strong preferences described by Eqs. (22)–(25) or (iii) pure non-traders described by Eqs. (8); CMM based on Eqs. (26)–(28);

	PDPM 1 (no RA)	PDPM 2 (no RA)	PDPM 1	PDPM 2
$\mathcal{L}(\hat{\beta})$	−4280.48	−4244.11	−4245.11	−4143.47

Each model specification was first estimated ignoring the panel nature of the data. The estimated parameters from these estimations were the starting values for the mixed model estimations on the longitudinal data. The models were estimated in *Biogeme* (Bierlaire, 2003) using 500 Halton draws. Table 6 summarizes the goodness of fit for the three models. Table 7 presents the estimation results. Random terms are estimated for three alternatives only. Theoretically, it would be possible to estimate random terms for all four alternatives, yet in practice, it was not possible to identify them all.

Both PDPMs 1 and 2 demonstrate a significant improvement in the goodness of fit in comparison with the BM, with PDPM 2 further increasing the likelihood. We have also estimated models PDPM 1 and PDPM 2 without the RA components. The log-likelihoods of these models, the detailed estimation results of which are not reported here, are respectively −4280.48 and −4244.11, indicating an improvement in performance both due to the additional class and the inclusion of the RA components, when considering the results in Table 6. All the estimated parameters of the class-specific model exhibit the expected signs and, with the exception of some alternative-specific constants and $\sigma^{\text{trader}}_{\text{WALK}}$ in PDPM 2, are significant. The results indicate strong inertia for bikers. Table A.1 displays the contributions of the travel time, ASCs and inertia, to the mode utilities of the second class, for indicative values of travel time (minimum and maximum), in order to allow an interpretation of the magnitudes of the parameters while taking into consideration the range of the attributes and the scale of the service attributes. The effect of inertia and the ASCs is not as strong as it initially reads when looking at the estimated parameters in Table 7. The results of PDPM 2 indicate high heterogeneity of preferences with respect to car ($\sigma^{\text{trader}}_{\text{CAR}} = 18.10$) in the first class and strong bias and lower heterogeneity of preferences with respect to bike.

The contextual parameter θ, associated with the RA component, plays an important role in the CMM; in particular in the distinction between the classes of utility maximizers and pure non-traders. Recall that the RA component is added in the class-membership function of utility-maximizers with strong preferences. In PDPM 1, with only two classes of respondents, θ takes a negative value, favoring the pure non-traders, while in PDPM 2, with three classes of respondents, it becomes positive, favoring, in line with our assumption, the class of utility-maximizers with strong preferences. The inclusion of the class of pure utility-maximizers (no inertia effects) appears to assist the model in discriminating between utility-maximizers with strong preferences and non-traders. It is interesting to observe the fluctuation of the class-membership components, denoted by π_d in Table 7, across the models. PDPM 1 estimates the share of utility-maximizing individuals to 75.5%, leaving the non-trading class with 24.5%. The latter is lower than the share of the respondents in the data that appear to be non-traders (55%) based on their choices across the 8 tasks. PDPM 2

[12] We have also tested the base model without inertia effects, i.e. with utility functions given by Eqs. (18)–(21), but it was inferior to the one including inertia effects. Therefore, it is assumed that inertia is present in the sample.

[13] We have also tested a PDPM with two classes of utility-maximizers, described respectively by Eqs. (18)–(21) and (22)–(25), but it was not possible to discriminate between the two in the absence of the pure non-trading class.

Table 7
Estimation results.

Parameter	BM	PDPM 1 Value (robust t-test)	PDPM 2
class-membership model			
ASC_{strong}	–	16.60 (2.36)	2.36 (3.00)
$ASC_{\text{non-trader}}$	–	0	−6.78 (−4.85)
θ	–	−7.74 (−2.02)	3.76 (6.38)
σ_{strong}	–	12.60 (2.26)	4.28 (7.46)
$\sigma_{\text{non-trader}}$	–	0	10.70 (6.80)
π_{trader}	**1**	–	**0.088**
π_{strong}	–	**0.755**	**0.771**
$\pi_{\text{non-trader}}$	–	**0.245**	**0.141**
class-specific model			
$ASC^{\text{trader}}_{\text{WALK}}$	–	–	4.26 (3.07)
$ASC^{\text{trader}}_{\text{BIKE}}$	–	–	13.1 (1.59)
$ASC^{\text{trader}}_{\text{CAR}}$	–	–	3.30 (0.49)
$ASC^{\text{strong}}_{\text{WALK}}$	1.15 (0.92)	−0.86 (−0.48)	1.89 (1.58)
$ASC^{\text{strong}}_{\text{BIKE}}$	−0.01 (−0.01)	−2.92 (−2.82)	−4.49 (−4.09)
$ASC^{\text{strong}}_{\text{CAR}}$	−0.82 (−3.74)	−1.20 (−3.35)	−0.78 (−2.36)
$\beta^{\leq 30\text{ min}}_{\text{walkTime}}$	−0.18 (−3.30)	−0.23 (−4.56)	−0.34 (−6.04)
$\beta^{>30\text{ min}}_{\text{walkTime}}$	−0.11 (−5.67)	−0.12 (−6.82)	−0.21 (−7.64)
$\beta_{\text{cycleTime}}$	−0.17 (−13.27)	−0.26 (−4.89)	−0.31 (−8.40)
$\beta_{\text{totalTime}}$	−0.12 (−13.90)	−0.12 (−10.92)	−0.23 (−11.47)
$\beta_{\text{totalCost}}$	−0.28 (−10.19)	−0.30 (−8.55)	−0.61 (−11.12)
$\beta^{\geq 2}_{\text{numTranfers}}$	−0.56 (−4.28)	−0.60 (−4.55)	−0.76 (−3.67)
ν_{walker}	4.10 (4.49)	4.13 (4.67)	3.44 (3.83)
ν_{cyclist}	9.06 (10.51)	6.07 (6.57)	10.09 (9.73)
ν_{driver}	2.57 (7.58)	2.18 (6.64)	2.65 (5.90)
ν_{PT}	2.76 (8.86)	1.23 (3.56)	3.03 (6.05)
$\sigma^{\text{trader}}_{\text{WALK}}$	–	–	−0.52 (−0.49)
$\sigma^{\text{trader}}_{\text{BIKE}}$	–	–	−1.80 (−7.65)
$\sigma^{\text{trader}}_{\text{CAR}}$	–	–	18.10 (5.16)
$\sigma^{\text{strong}}_{\text{WALK}}$	3.43 (8.71)	4.21 (6.73)	5.39 (6.46)
$\sigma^{\text{strong}}_{\text{BIKE}}$	4.66 (13.06)	4.56 (9.23)	5.61 (8.45)
$\sigma^{\text{strong}}_{\text{CAR}}$	3.05 (15.04)	2.66 (10.95)	3.28 (10.77)
VoT (CHF/h)	**24.9**	**18.5**	**19.4**

decreases the share of non-traders to 14.1%, with 8.8% now representing the pure utility-maximizers and 77.1% those with strong preferences.

Fig. 3 illustrates the estimated persistence by PDPM 2 for trading and non-trading respondents respectively. This is computed as the average of the choice probability of the chosen alternative over the choice tasks, when this corresponds to the RP choice of the respondent. Comparing the estimated persistence of trading and non-trading respondents, we observe that, according to the expectations, the one concerning the latter group is higher. Fig. 4 illustrates the effect of the normalized overall RA on the choice probability. As it can be seen in the plot, as well as in the derivations of the mode choice probability with respect to the RA component in Appendix B, we cannot know the net effect of the RA on the choice probability. It may be positive or negative, depending on the magnitudes of the choice probabilities [see Eq. (33) of the Appendix].

Fig. 5 illustrates the distribution of the normalized overall RA for PDPM 2, including a segmentation by trading/non-trading respondents (as observed in the data). The figure confirms the assumption of the model that certain respondents appear to be non-traders as their chosen alternative is typically more attractive in the experiment in comparison with the remaining alternatives. Based on the data 832 (55%) respondents can be classified as non-traders; given that they systematically chose their RP choice across all 8 tasks. According to the final model (PDPM 2), approximately 215 respondents (∼14%) are non-traders. This is in line with the assumption of the modeling framework, i.e. that certain utility-maximizing individuals may exhibit non-trading behavior as a result of the combination of (i) their strong preference for a specific mode and (ii) the relative advantage of their preferred mode over the remaining alternatives in the choice experiment. Along these lines, we expect the model to predict a lower share of non-trading respondents and distinguish between pure non-trades and utility-maximizers with strong preferences. Fig. 5 supports this assumption. We observe that for respondents that, according to the data, are non-traders, the distribution of the RA component is shifted to the right, while for trading respondents it is more uniformly distributed. In particular, for about 40% of the observations pertaining to non-traders the normalized overall RA is greater than 0.9, signifying that the preferred mode is in an advantageous position in comparison with the remaining alternatives.

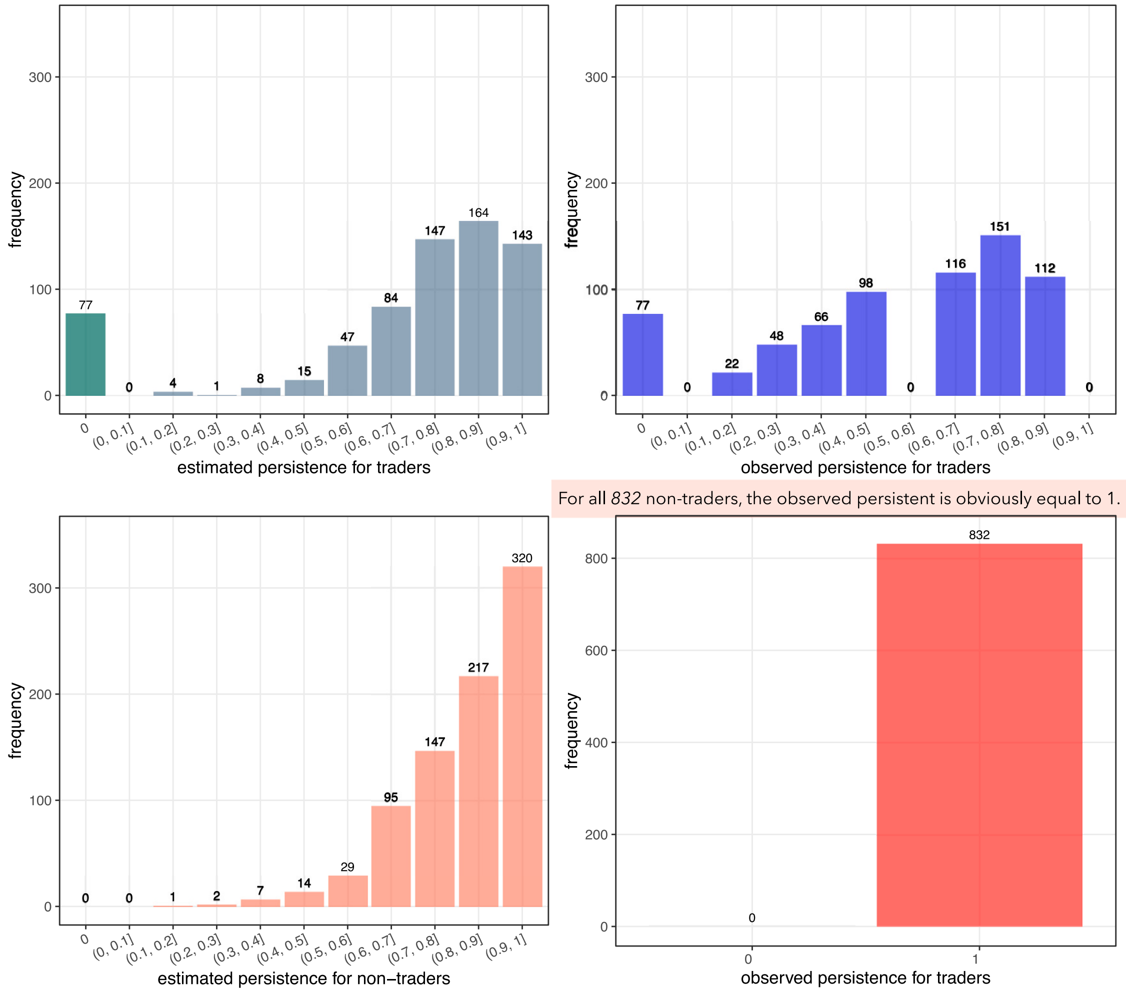

Fig. 3. Distribution of the estimated $\left(\frac{\sum_t \Pr_{nt}(i)\times I_p}{8}, \text{ where } I_p = 1 \text{ if } i = p, \text{ and } 0 \text{ otherwise.}\right)$ and observed persistence for trading and non-trading respondents.

Table 8 provides a summary of descriptive statistics about the overall RA and RD components. Same as Fig. 5, Table 8 illustrates that for the segment of respondents exhibiting non-trading behavior in the experiment – in contrast with the segment of trading respondents – the RP mode is in more advantageous positions, as to total time and cost, with respect to the remaining alternatives.[14]

4.4. Value of time

The value of time (VoT) for the proposed framework is computed as follows

$$VoT_n = \sum_{d\in D} \Pr_{nt}(d) \frac{\partial U^d/\partial Time}{\partial U^d/\partial Cost} = \sum_{d\in D} \Pr_{nt}(d) \frac{\beta^d_{\text{totalTime}}}{\beta^d_{\text{totalCost}}}. \tag{29}$$

while for the average individual in BM, it is simply

$$VoT = \frac{\partial U/\partial Time}{\partial U/\partial Cost} = \frac{\beta_{\text{totalTime}}}{\beta_{\text{totalCost}}}. \tag{30}$$

Table 9 presents basic statistics on the VoT as derived from the three models. Fig. 6 depicts the distribution of the VoT for each of the PDPMs. PDPMs 1 and 2 reduce the mean VoT predicted by the BM by 6.4 CHF/h and 5.5 CHF/h, respectively. As

[14] *Remark*: Only in 4% of the choice tasks the RP mode corresponds to a dominant option with respect to time and cost.

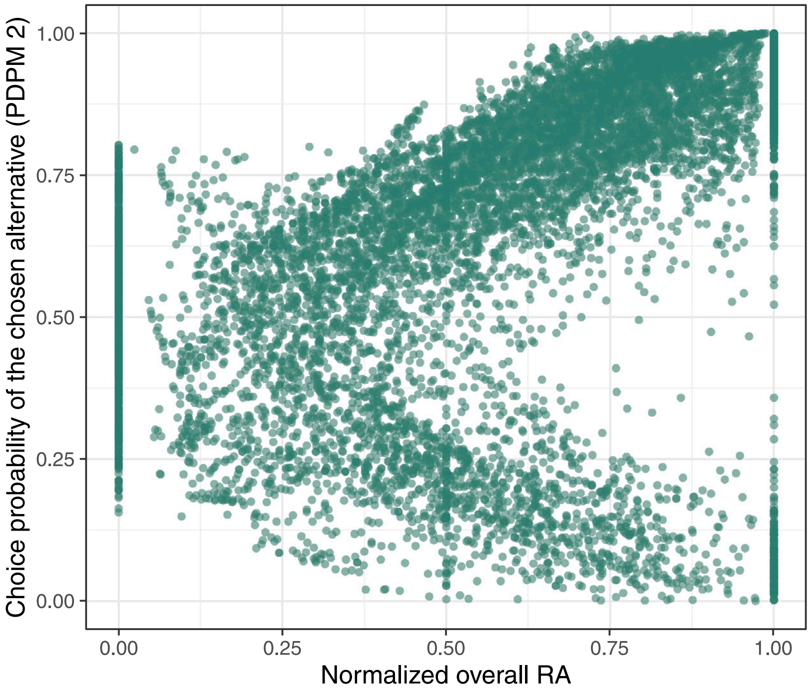

Fig. 4. Scatter plot of the choice probability of the chosen alternative against the normalized overall RA $\frac{\sum_j RA(p,j)}{\sum_j RA(p,j)+\sum_j RD(p,j)}$.

Table 8
RA component statistics.

Trading respondents	PDMP 1		PDMP 2	
	$\sum_j RA(p,j)$	$\sum_j RD(p,j)$	$\sum_j RA(p,j)$	$\sum_j RD(p,j)$
min	0	0	0	0
max	2	2	2	2
mean	0.844	0.677	0.849	0.672
stdev	0.502	0.488	0.520	0.504
Non-trading respondents				
min	0	0	0	0
max	2	2	2	2
mean	1.053	0.433	1.065	0.421
stdev	0.462	0.460	0.474	0.468

expected, the VoT distribution of the PDPM 1, which entails a higher share of non-traders with $VoT = 0$ CHF/h, is shifted towards a lower range of values, and is characterized by bigger standard deviation, in comparison with the distribution derived from the PDPM 2.

5. Conclusion

We have presented a discrete choice modeling framework with heterogeneous decision rules accounting for non-trading behavior and its application to a Swiss SP mode choice dataset. The approach incorporates inertia effects and tackles serial correlation in the SP data. It is the first to consider contextual effects on the probability of a decision rule being employed by an individual; this is accomplished through the inclusion of a relative advantage component in the class-membership model. The modeling framework discriminates between utility-maximizing individuals with strong preferences and pure non-trading individuals, both of whom

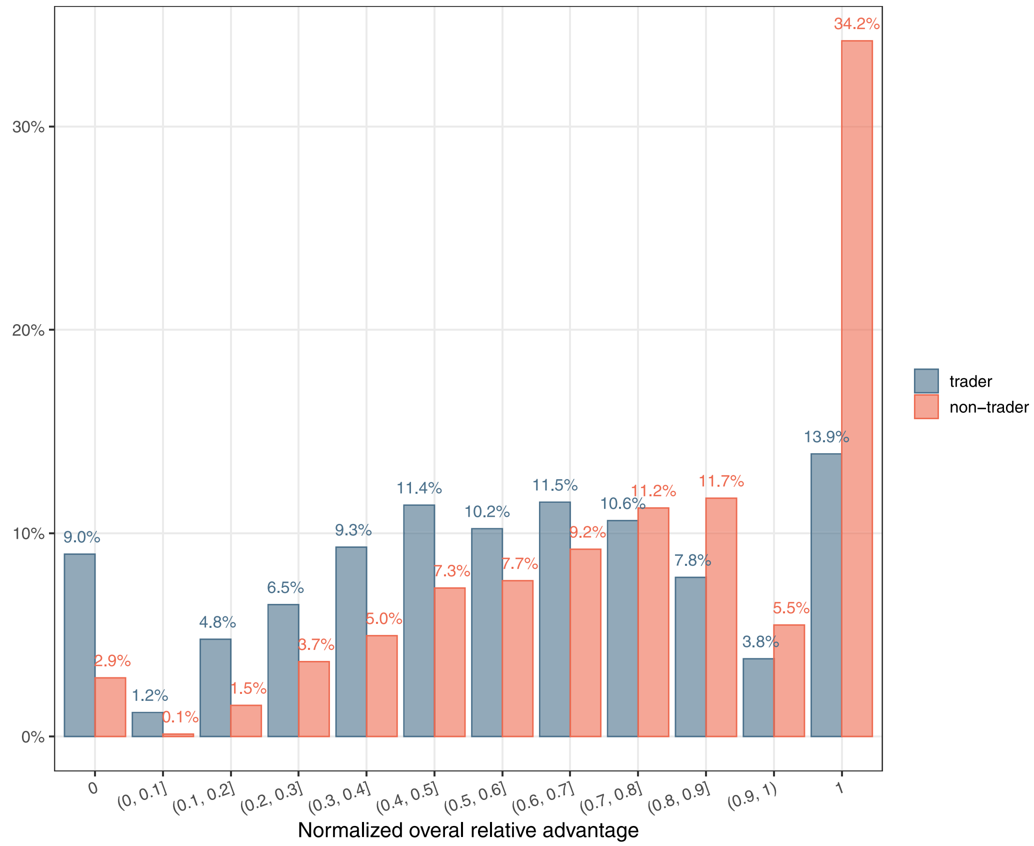

Fig. 5. Distribution of the normalized overall RA $\frac{\sum_j RA(p,j)}{\sum_j RA(p,j)+\sum_j RD(p,j)}$ per respondent segment (trading versus non-trading individuals as observed in the data) for PDPM 2. A value of 1 represents the cases where ones preferred (RP) mode p is in an advantageous position, either by being (i) the fastest and cheapest option, (ii) the fastest yet not the most expensive or (iii) the cheapest yet not the slowest option in the choice task—as opposed to a value of 0 that represents the cases where p is either (i) the slowest and most expensive option in the task, (ii) the slowest yet not the cheapest or (iii) the most expensive yet not the fastest option. A value of 0.5 signifies that p is either the fastest and most expensive, or the slowest and cheapest option in the task. The remaining values depend on the estimated parameters and the position of p in comparison with the remaining alternatives in the task.

Table 9
Value of time statistics.

CHF/h	BM	PDPM 1	PDPM 2
min	–	12.9	17.6
max	–	22.3	21.2
mean	**24.9**	**18.5**	**19.4**
stdev	–	2.3	0.7

are expected to exhibit non-trading behavior. As expected, accounting for non-trading behavior results in important shifts in the estimated values of time.

The model specification described in this paper can be extended to further account for inter-personal heterogeneity, through the inclusion of random parameters β_n, associated with the attributes of the alternatives, as well as the inertia effects, in the CSM, and random parameters θ_n, associated with the contextual effects in the CMM. We plan to extend the model with a full specification of taste heterogeneity, in order to additionally investigate the expected confounding effects between heterogeneity in the decision-making process and 'common' taste heterogeneity (see Hess et al., 2012; Balbontin et al., 2019). Furthermore, we recognize the limitation of the current specification due to the constrained coefficients between the two classes of utility maximizers. Further investigation into the assumption of equal direct sensitivities to travel attributes for these two classes is needed. Future research will consider different coefficients across class-membership and individuals in order to better measure contextual inertia effects.

Finally, it is of interest to apply the proposed framework to other case studies and choice contexts, in order to investigate the presence of common patterns among datasets, as well as the ability of the model to disentangle the different manifestations of non-trading behavior.

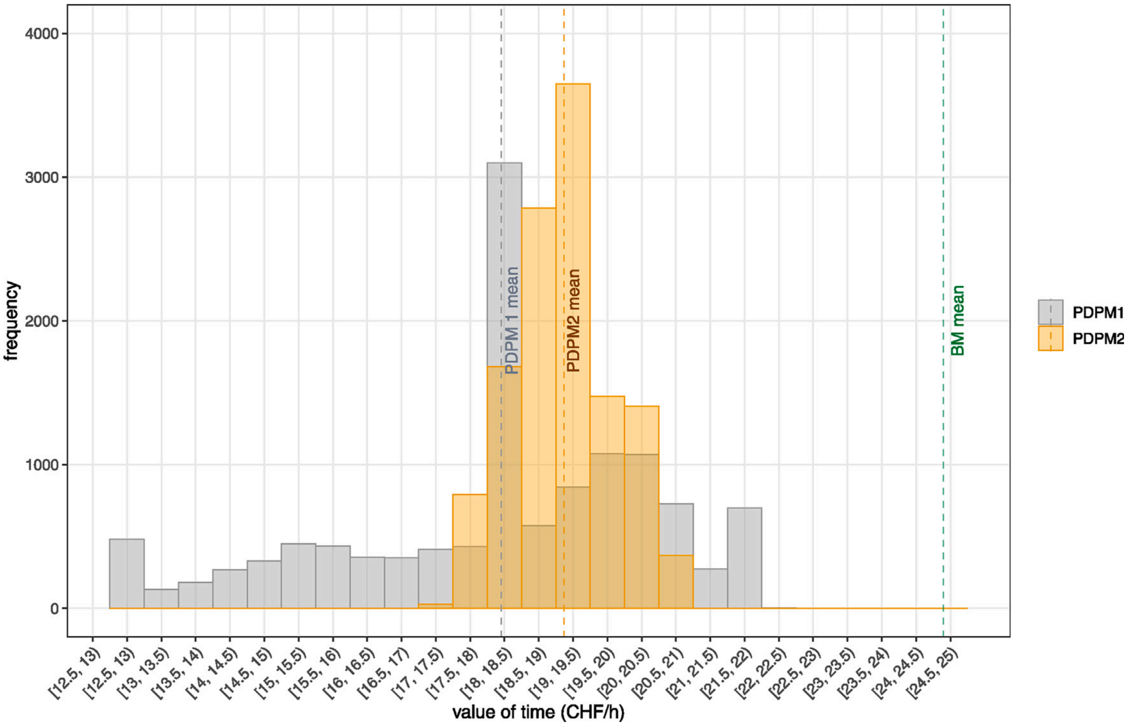

Fig. 6. Distribution of the value of time for the PDPMs.

CRediT authorship contribution statement

Evanthia Kazagli: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Matthieu de Lapparent:** Funding acquisition, Methodology, Project administration, Resources, Writing – review & editing.

Declaration of competing interest

None

Acknowledgments

This research is supported by the Swiss National Science Foundation Spark Grant **#CRSK-1_190745** *"Integrated Decision Heuristics & Discrete Choice Modeling: An Operational Framework for Demand Analysis"*. The authors would like to thank Antonin Danalet and the Federal Office for Spatial Development ARE for the provision of the data.

Appendix A

See Table A.1.

Table A.1
Contribution of travel time, *ASCs* and inertia on the overall utility.

Mode	Total time [minimum]	Total time [maximum]	Beta time	Total time min contr.	Total time max contr.	*ASC* contr.	Inertia max cont.
Walk	6	467	−0.34	−2.04	−158.78	1.89	3.44
Bike	2	140	−0.31	−0.62	−43.4	−4.49	10.09
Car	2	395	−0.23	−0.46	−90.85	−0.78	2.65
pt	2	1004	−0.23	−0.46	−230.92	0	3.03

Appendix B. Derivation of the mode choice probability with respect to the RA component

Derivation of the mode choice probability with respect to the RA component in order to illustrate the effect of its inclusion (in the class-membership model) on the choice probability. The choice probability can be expressed as $\mathrm{Pr}_i(RA) = \sum_d \mathrm{Pr}_{i|d}(RA)\cdot \mathrm{Pr}_d(RA)$. The derivative with respect to the RA component is

$$\frac{\partial \mathrm{Pr}_i}{\partial RA_c} = \sum_d \left(\underbrace{\frac{\partial \mathrm{Pr}_{i|d}(RA_c)}{\partial RA_c}}_{=0 \forall d, V_i} \cdot \underbrace{\mathrm{Pr}_d(RA_c)}_{>0} + \underbrace{\frac{\partial \mathrm{Pr}_d(RA_c)}{\partial RA_c}}_{>0 \text{ for } d=2;\, <0 \text{ for } d\neq 2} \cdot \underbrace{\mathrm{Pr}_{i|d}(RA_c)}_{>0} \right). \tag{31}$$

The RA component is only included in the class-membership function of the second class (utility-maximizers with strong preferences). As a result the derivative in Eq. (2) is 0 $\forall d \neq 2$. Then,

$$\begin{aligned}
\frac{\partial \mathrm{Pr}_i}{\partial RA_2} &= \sum_{d=1}^{3} \frac{\partial \mathrm{Pr}_d(RA_2)}{\partial RA_2} \cdot \mathrm{Pr}(i \mid d) = \\
&- \theta \cdot \mathrm{Pr}(d=1) \cdot \mathrm{Pr}(d=2) \cdot \mathrm{Pr}(i \mid 1) + \\
&+ \theta \cdot \mathrm{Pr}(d=2) \cdot [1 - \mathrm{Pr}(d=2)] \cdot \mathrm{Pr}(i \mid 2) - \\
&- \theta \cdot \mathrm{Pr}(d=3) \cdot \mathrm{Pr}(d=2) \cdot \mathrm{Pr}(i \mid 3) = \\
&= \theta \cdot \mathrm{Pr}(d=2) \cdot [-\mathrm{Pr}(d=1) \cdot \mathrm{Pr}(i \mid 1) + \\
&+ [1 - \mathrm{Pr}(d=2)] \cdot \mathrm{Pr}(i \mid 2) - \mathrm{Pr}(d=3) \cdot \mathrm{Pr}(i \mid 3)],
\end{aligned} \tag{32}$$

where $\mathrm{Pr}_c = \frac{e^{F_c}}{\sum_{c\neq d} e^{F_c} + e^{F_d + \theta \cdot RA}} \Rightarrow \frac{\partial \mathrm{Pr}_c}{\partial RA_2} = \begin{cases} \theta \cdot \mathrm{Pr}_d(1 - \mathrm{Pr}_d), & \text{for } c = d, \text{ and} \\ -\theta \cdot \mathrm{Pr}_c \cdot \mathrm{Pr}_d, & \text{for } c \neq d. \end{cases}$

Eq. (32) can also be expressed as

$$\begin{aligned}
\frac{\partial \mathrm{Pr}_i}{\partial RA_2} &= \theta \cdot \mathrm{Pr}(d=2) \cdot \left(\mathrm{Pr}(i \mid 2) - \underbrace{\sum_{d=1}^{3} \mathrm{Pr}(i \mid d) \cdot \mathrm{Pr}(d)}_{\mathrm{Pr}(i)} \right) \Rightarrow \\
\frac{\partial \mathrm{Pr}_i}{\partial RA_2} &= \theta \cdot \mathrm{Pr}(d=2) \cdot (\mathrm{Pr}(i \mid 2) - \mathrm{Pr}(i)) \text{ and is } \begin{cases} >0 & \text{if } \mathrm{Pr}(i \mid 2) > \mathrm{Pr}(i) \\ =0 & \text{if } \mathrm{Pr}(i \mid 2) = \mathrm{Pr}(i) \\ <0 & \text{if } \mathrm{Pr}(i \mid 2) < \mathrm{Pr}(i) \end{cases}
\end{aligned} \tag{33}$$

References

Adamowicz, W.L., Swait, J.D., 2012. Are food choices really habitual? Integrating habits, variety-seeking, and compensatory choice in a utility-maximizing framework. Am. J. Agric. Econ. 95 (1), 17–41, URL: https://doi.org/10.1093/ajae/aas078.

Balbontin, C., Hensher, D.A., Collins, A.T., 2017. Integrating attribute non-attendance and value learning with risk attitudes and perceptual conditioning. Transp. Res. E 97, 172–191, URL: http://www.sciencedirect.com/science/article/pii/S1366554516305178.

Balbontin, C., Hensher, D.A., Collins, A.T., 2019. How to better represent preferences in choice models: The contributions to preference heterogeneity attributable to the presence of process heterogeneity. Transp. Res. B 122, 218–248, URL: http://www.sciencedirect.com/science/article/pii/S0191261518307859.

Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models.

Boeri, M., Scarpa, R., Chorus, C.G., 2014. Stated choices and benefit estimates in the context of traffic calming schemes: Utility maximization, regret minimization, or both? Transp. Res. A 61, 121–135, URL: http://www.sciencedirect.com/science/article/pii/S0965856414000044.

Bradley, M., Daly, a., Stopher, P., Lee-Gosselin, M., 1996. Estimation of logit choice models using mixed stated-preference and revealed-preference information. pp. 209–231.

Cantillo, V., Ortúzar, J.d.D., Williams, H.C.W.L., 2007. Modeling discrete choices in the presence of inertia and serial correlation. Transp. Sci. 41 (2), 195–205, URL: https://pubsonline.informs.org/doi/abs/10.1287/trsc.1060.0178.

Cherchi, E., Cirillo, C., 2014. Understanding variability, habit and the effect of long period activity plan in modal choices: a day to day, week to week analysis on panel data. Transportation 41 (6), 1245–1262, URL: https://ideas.repec.org/a/kap/transp/v41y2014i6p1245-1262.html.

Cherchi, E., Manca, F., 2011. Accounting for inertia in modal choices: some new evidence using a RP/SP dataset. Transportation 38 (4), 679, URL: https://doi.org/10.1007/s11116-011-9338-9.

Dey, B.K., Anowar, S., Eluru, N., Hatzopoulou, M., 2018. Accommodating exogenous variable and decision rule heterogeneity in discrete choice models: Application to bicyclist route choice. PLoS One 13 (11), 1–19, URL: https://doi.org/10.1371/journal.pone.0208309.

Elrod, T., Johnson, R.D., White, J., 2004. A new integrated model of noncompensatory and compensatory decision strategies. Organ. Behav. Hum. Decis. Process. 95 (1), 1–19, URL: http://www.sciencedirect.com/science/article/pii/S0749597804000573.

Gärling, T., Axhausen, K.W., 2003. Introduction: Habitual travel choice. Transportation 30 (1), 1–11, URL: https://doi.org/10.1023/A:1021230223001.

Hensher, D.A., 2019. Context dependent process heuristics and choice analysis – A note on two interacting themes linked to behavioural realism. Transp. Res. A 125, 119–122, URL: http://www.sciencedirect.com/science/article/pii/S0965856419301107.

Hensher, D.A., Balbontin, C., Collins, A.T., 2018. Heterogeneity in decision processes: Embedding extremeness aversion, risk attitude and perceptual conditioning in multiple process rules choice making. Transp. Res. A 111, 316–325, URL: http://www.sciencedirect.com/science/article/pii/S096585641730397X.

Hensher, D.A., Collins, a.T., Greene, W.H., 2013. Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding. Transportation 40 (5), 1003–1020, URL: https://doi.org/10.1007/s11116-012-9447-0.

Hensher, D.A., Greene, W.H., 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. Empir. Econ. 39 (2), 413–426, URL: https://doi.org/10.1007/s00181-009-0310-x.

Hess, S., Rose, J.M., Polak, J., 2010. Non-trading, lexicographic and inconsistent behaviour in stated choice data. Transp. Res. D 15 (7), 405–417, URL: http://www.sciencedirect.com/science/article/pii/S1361920910000684. Specification and interpretation issues in behavioural models used for environmental assessment.

Hess, S., Stathopoulos, A., 2013. A mixed random utility — Random regret model linking the choice of decision rule to latent character traits. J. Choice Model. 9, 27–38, URL: http://www.sciencedirect.com/science/article/pii/S1755534513000626. Issues in Choice Modelling: selected papers from the 13th International Conference on Travel Behaviour Research.

Hess, S., Stathopoulos, A., Daly, a., 2012. Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. Transportation 39 (3), 565–591, URL: https://doi.org/10.1007/s11116-011-9365-6.

Kivetz, R., Netzer, O., Srinivasan, V., 2004. Alternative models for capturing the compromise effect. J. Mar. Res. 41 (3), 237–257, URL: https://doi.org/10.1509/jmkr.41.3.237.35990.

Leong, W., Hensher, D.A., 2012a. Embedding decision heuristics in discrete choice models: A review. Transp. Rev. 32 (3), 313–331, https://doi.org/10.1080/01441647.2012.671195.

Leong, W., Hensher, D.A., 2012b. Embedding multiple heuristics into choice models: an exploratory analysis. J. Choice Model. 5 (3), 131 – 144, URL: http://www.sciencedirect.com/science/article/pii/S175553451300002X.

Leong, W., Hensher, D.A., 2014. Relative advantage maximisation as a model of context dependence for binary choice data. J. Choice Model. 11, 30–42, URL: http://www.sciencedirect.com/science/article/pii/S1755534514000256. Process heuristics in choice analysis.

McAlister, L., Srivastava, R., Horowitz, J., Jones, M., Kamakura, W., Kulchitsky, J., Ratchford, B., Russel, G., Sultan, F., Yai, T., Weiss, D., Winer, R., 1991. Incorporating choice dynamics in models of consumer behavior. Mark. Lett. 2 (3), 241–252, URL: https://doi.org/10.1007/BF00554129.

McNair, B.J., Hensher, D.A., Bennett, J., 2012. Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: A probabilistic decision process model. Environ. Resour. Econ. 51 (4), 599–616, URL: https://doi.org/10.1007/s10640-011-9514-6.

Schmid, B., Jokubauskaite, S., Aschauer, F., Peer, S., Hössinger, R., Gerike, R., Jara-Diaz, S.R., Axhausen, K.W., 2019. A pooled RP/SP mode, route and destination choice model to investigate mode and user-type effects in the value of travel time savings. Transp. Res. A 124, 262–294, URL: https://www.sciencedirect.com/science/article/pii/S0965856418301721.

Train, K.E., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press.

Tversky, A., Simonson, I., 1993. Context-dependent preferences. Manage. Sci. 39 (10), 1179–1189.

Weis, C., Kowald, M., Danalet, A., Schmid, B., Vrtic, M., Axhausen, K.W., Mathys, N., 2021. Surveying and analysing mode and route choices in Switzerland 2010–2015. Travel Behav. Soc. 22, 10–21, URL: https://www.sciencedirect.com/science/article/pii/S2214367X20302003.

Zhu, W., Timmermans, H., 2010. Cognitive process model of individual choice behaviour incorporating principles of bounded rationality and heterogeneous decision heuristics. Environ. Plan. B: Plann. Des. 37 (1), 59–74, URL: https://doi.org/10.1068/b35039.